WHAT IS CLAIMED IS:

1. A recombinant, purified or isolated polynucleotide comprising a mammalian PG1 gene, cDNA, complement thereof, or fragment thereof having at least 10 nucleotides in length.

2. The polynucleotide according to claim 1, wherein said mammalian PG1 gene or cDNA is human or mouse.

3. The polynucleotide according to claim 2, wherein the polynucleotide is selected from SEQ ID NOs: 3, 69, 112-124, 179, and 182-184.

4. A polynucleotide selected from SEQ ID NOs: 185-578.

5. A purified or isolated polypeptide comprising a mammalian PG1 protein, or fragment thereof having at least 8 amino acids in length.

6. The polypeptide according to claim 5, wherein said mammalian PG1 protein is human or mouse.

7. The polypeptide according to claim 6, wherein said polypeptide is selected from SEQ ID NOs: 4, 5, 70, 74, and 125-136.

8. The polypeptide according to claim 5, wherein said polypeptide consists of said mammalian PG1 protein, or fragment thereof having at least 8 amino acids in length.

9. A polynucleotide comprising a nucleic acid sequence encoding a polypeptide according to claim 8.

10. An antibody composition capable of selectively binding to an epitope-containing fragment of a polypeptide according to claim 8, wherein said antibody is either polyclonal or monoclonal.

11. A vector comprising a polynucleotide according to any one of claims 1, 4, and 9.

12. A host cell comprising a polynucleotide according to claim 11.

13 A nonhuman host animal or mammal comprising a vector according to claim 11.

14. A mammalian host cell comprising a PG1 gene disrupted by homologous recombination with a knock out vector.

15. A nonhuman host mammal comprising a PG1 gene disrupted by homologous recombination with a knock out vector.

16. A polynucleotide according to any one of claims 1, 4, and 9, further comprising a label.

17. A polynucleotide according to any one of claims 1, 4, and 9, attached to a solid support.

18. A random or addressable array of polynucleotides comprising at least one polynucleotide according to any one of claims 1, 4, and 9.

19. A method of determining whether an individual is at risk of developing cancer or prostate cancer, or whether said individual suffers from cancer or prostate cancer as a result of a mutation in the PG1 gene comprising:

obtaining a nucleic acid sample from said individual; and

determining whether the nucleotides present at one or more PG1-related biallelic marker are indicative of a risk of developing cancer or prostate cancer or indicative of cancer or prostate cancer resulting from a mutation in the PG1 gene.

20. A method of determining whether an individual is at risk of developing cancer or prostate cancer or whether said individual suffers from cancer or prostate cancer as a result of a mutation in the PG1 gene comprising:

obtaining a nucleic acid sample from said individual; and

determining whether the nucleotides present at one or more PG1-related biallelic marker are indicative of a risk of developing cancer or prostate cancer or indicative of cancer or prostate cancer resulting from a mutation in the PG1 gene.

21. A method according to either one of claims 19 and 20, wherein said PG1-related biallelic is a PG1-related biallelic marker positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576 , and 4-66.

22. A method of obtaining an allele of the PG1 gene which is associated with a detectable phenotype comprising:

obtaining a nucleic acid sample from an individual expressing said detectable phenotype;

contacting said nucleic acid sample with an agent capable of specifically detecting a nucleic acid encoding the PG1 protein; and

isolating said nucleic acid encoding the PG1 protein.

23. A method of obtaining an allele of the PG1 gene which is associated with a detectable phenotype comprising:

obtaining a nucleic acid sample from an individual expressing said detectable phenotype;

contacting said nucleic acid sample with an agent capable of specifically detecting a sequence within the 8p23 region of the human genome;

identifying a nucleic acid encoding the PG1 protein in said nucleic acid sample; and

isolating said nucleic acid encoding the PG1 protein.

24. A method of categorizing the risk of prostate cancer in an individual comprising the step of assaying a sample taken from the individual to determine whether the individual carries an allelic variant of PG1 associated with an increased risk of prostate cancer.

25. The method of Claim 24 wherein said sample is a nucleic acid sample.

26. The method of Claim 24 wherein said sample is a protein sample.

27. The method of Claim 26, further comprising determining whether the PG1 protein in said sample binds an antibody that binds specifically to a PG1 isoform associated with prostate cancer.

28. A method of genotyping comprising determining the identity of a nucleotide at a PG1-related biallelic marker in a biological sample.

29. A method of estimating the frequency of an allele in a population comprising determining the proportional representation of a nucleotide at a PG1-related biallelic marker in a pooled biological sample derived from said population.

30. A method of detecting an association between a genotype and a phenotype, comprising the steps of:

a) genotyping at least one PG1-related biallelic marker in a trait positive population;

b) genotyping said PG1-related biallelic marker in a control population; and

c) determining whether a statistically significant association exists between said genotype and said phenotype.

31. A method of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising:

a) genotyping at least one PG1-related biallelic marker;

b) genotyping a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome of each individual in said population; and

c) applying an haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency.

32. A method of detecting an association between a haplotype and a phenotype, comprising the steps of:

a) estimating the frequency of at least one haplotype in a trait positive population according to the method of claim 31;

b) estimating the frequency of said haplotype in a control population according to the method of claim 31; and

c) determining whether a statistically significant association exists between said haplotype and said phenotype.

33. A method according to claim 31, wherein said PG1-related biallelic marker and said second biallelic marker are 4-77/151 and 4-66/145,

34. A method according to claim 32, wherein said haplotype exhibits a p-value of < 1x $10^{-3}$ in an association with a trait positive population with cancer, or prostate cancer.

35. A method according to any one of claims 29 to 31, wherein said PG1-related biallelic is a PG1-related biallelic marker positioned in SEQ ID NO: 179; a PG1-related biallelic marker selected from the group consisting of 99-1485/251, 99-622/95, 99-619/141, 4-76/222, 4-77/151, 4-71/233, 4-72/127, 4-73/134, 99-610/250, 99-609/225, 4-90/283, 99-602/258, 99-600/492, 99-598/130, 99-217/277, 99-576/421, 4-61/269, 4-66/145, and 4-67/40; or a PG1-related biallelic marker selected from the group consisting of 99-622, 4-77, 4-71, 4-73, 99-598, 99-576 , and 4-66.

36. A method according to either one of claims 30 and 32, wherein said control population is a trait negative population or a random population.

37. A method according to any one of claims 22, 23, 30, and 32, wherein said phenotype is a disease, cancer or prostate cancer; a response to an anti-cancer agent or an anti-prostate cancer agent; or a side effect to an anti-cancer or anti-prostate cancer agent.

38. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 179 or the complements thereof, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 179: 1-2324, 2852-2936, 3204-3249, 3456-3572, 3899-4996, 5028-6086, 6310-8710, 9136-11170, 11534-12104, 12733-13163, 13206-14150, 14191-14302, 14338-14359, 14788-15589, 16050-16409, 16440-21718, 21959-22007, 22086-23057, 23488-23712, 23832-24099, 24165-24376,

24429-24568, 24607-25096, 25127-25269, 25300-27576, 27612-29217, 29415-30776, 30807-30986, 31628-32658, 32699-36324, 36772-39149, 39184-40269, 40580-40683, 40844-41048, 41271-43539, 43570-47024, 47510-48065, 48192-49692, 49723-50174, 52626-53599, 54516-55209, and 55666-56146.

39. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 3: 1-280, 651-690, 3315-4288, and 5176-5227.

40. An isolated, purified, or recombinant polynucleotide which encodes a polypeptide comprising a contiguous span of at least 8 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1 of the amino acid positions 1-26, 295-302, and 333-353

41. An isolated, purified, or recombinant polypeptide comprising a contiguous span of at least 8 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1 of the amino acid positions 1-26, 295-302, and 333-353

42. An isolated or purified antibody composition are capable of selectively binding to an epitope-containing fragment of a polypeptide according to claim 55, wherein said epitope comprises at least 1 of the amino acid positions 1-26, 295-302, and 333-353

43. A computer readable medium having stored thereon a sequence selected from the group consisting of a nucleic acid code comprising one of the following:

a) a contiguous span of at least 12 nucleotides of SEQ ID No 179, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 179: 1-2324, 2852-2936, 3204-3249, 3456-3572, 3899-4996, 5028-6086, 6310-8710, 9136-11170, 11534-12104, 12733-13163, 13206-14150, 14191-14302, 14338-14359, 14788-15589, 16050-16409, 16440-21718, 21959-22007, 22086-23057, 23488-23712, 23832-24099, 24165-24376, 24429-24568, 24607-25096, 25127-25269, 25300-27576, 27612-29217, 29415-30776, 30807-30986, 31628-32658, 32699-36324, 36772-39149, 39184-40269, 40580-40683, 40844-41048, 41271-43539, 43570-47024, 47510-48065, 48192-49692, 49723-50174, 52626-53599, 54516-55209, and 55666-56146;

b) a contiguous span of at least 12 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 3: 1-280, 651-690, 3315-4288, and 5176-5227; and

c) a nucleotide sequence complementary to either one of the preceding nucleotide sequences.

44. A computer readable medium having stored thereon a sequence consisting of a polypeptide code comprising a contiguous span of at least 8 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1 of the amino acid positions 1-26, 295-302, and 333-353.

45. A computer system comprising a processor and a data storage device wherein said data storage device a computer readable medium according to with claim 43 or 44.

46. A computer system according to claim 45, further comprising a sequence comparer and a data storage device having reference sequences stored thereon.

47. A computer system of Claim 46 wherein said sequence comparer comprises a computer program which indicates polymorphisms.

48. A computer system of Claim 45 further comprising an identifier which identifies features in said sequence.

49. A method for comparing a first sequence to a reference sequence, comprising the steps of:

reading said first sequence and said reference sequence through use of a computer program which compares sequences; and

determining differences between said first sequence and said reference sequence with said computer program,

wherein said first sequence is selected from the group consisting of a nucleic acid code comprising one of the following:

a) a contiguous span of at least 12 nucleotides of SEQ ID No 179, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 179: 1-2324, 2852-2936, 3204-3249, 3456-3572, 3899-4996, 5028-6086, 6310-8710, 9136-11170,

11534-12104, 12733-13163, 13206-14150, 14191-14302, 14338-14359, 14788-15589, 16050-16409, 16440-21718, 21959-22007, 22086-23057, 23488-23712, 23832-24099, 24165-24376, 24429-24568, 24607-25096, 25127-25269, 25300-27576, 27612-29217, 29415-30776, 30807-30986, 31628-32658, 32699-36324, 36772-39149, 39184-40269, 40580-40683, 40844-41048, 41271-43539, 43570-47024, 47510-48065, 48192-49692, 49723-50174, 52626-53599, 54516-55209, and 55666-56146;

b) a contiguous span of at least 12 nucleotides of SEQ ID No 3 or the complements thereof, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 3: 1-280, 651-690, 3315-4288, and 5176-5227;

c) a nucleotide sequence complementary to either one of the preceding nucleotide sequences; and

d) a polypeptide code comprising a contiguous span of at least 8 amino acids of SEQ ID No 4, wherein said contiguous span includes at least 1 of the amino acid positions 1-26, 295-302, and 333-353.